E7.6-10.0.0.2
CR-119104

(E76-10002) CANONICAL ANALYSIS AND
TRANSFORMATION OF SKYLAB MULTISPECTRAL
SCANNER DATA Interim Report (Pennsylvania
State Univ.) 6 p HC $3.25 CSCL 08L

N76-10534

Unclas
G3/43 00002

Interim Report

ORSER-SSEL Technical Report 18-75

CANONICAL ANALYSIS AND TRANSFORMATION OF SKYLAB MULTISPECTRAL
SCANNER DATA

D. M. Barr and B. F. Merembeck

EREP Investigation 475
Contract Number NAS 9-13406

INTERDISCIPLINARY APPLICATIONS AND INTERPRETATIONS OF EREP DATA
WITHIN THE SUSQUEHANNA RIVER BASIN

Resource Inventory, Land Use, and Pollution

Office for Remote Sensing of Earth Resources (ORSER)
Space Science and Engineering Laboratory (SSEL)
Room 219 Electrical Engineering West
The Pennsylvania State University
University Park, Pennsylvania 16802

Principal Investigators:

Dr. George J. McMurtry
Dr. Gary W. Petersen

Date: August 1975

---

*USER IDENTIFIED SIGNIFICANT RESULTS

ABSTRACT

ORSER-SSEL Technical Report 18-75

CANONICAL ANALYSIS AND TRANSFORMATION OF SKYLAB MULTISPECTRAL SCANNER DATA
D. M. Barr and B. F. Merembeck

Initial work with digital data from selected Skylab scenes resulted in the selection of sixteen channels (eleven bands) and in the development of 22 signatures. The scenes were all from SL3, Orbit 14, 5 August 1973, Tape 933847. The objective in this study was to process the data by means of canonical analysis, with a view toward improvement of classification, reduction of data bulk, and determination of the value of each band for classification of various targets.

Statistical information for each of the 22 signatures was obtained from the STATS program, and training areas for specific targets were defined. The new signatures, number of observations, and variance-covariance matrixes for the signatures were obtained and used as input to the CANAL program. The objective here was to maximize the separability of the categories (the transformation will usually also reduce the within-category variance).

It was determined that the first three axes contained 98.83% of the variance contained within the transformation; hence, only the first three axes were used in the transformation. Thus, canonical transformation is a form of feature selection, resulting here in an 81.25% reduction in data bulk.

Evaluation of the classification, now possible using the transformed data, has begun.

CANONICAL ANALYSIS AND TRANSFORMATION

OF SKYLAB MULTISPECTRAL SCANNER DATA

D. M. Barr and B. F. Merembeck

OBJECTIVES

The initial results of channel evaluation and classification of selected Skylab scenes (ORSER-SSEL Technical Report 2-75) resulted in the selection of sixteen channels (eleven bands) of data and development of twenty-two spectral signatures (six categories). All scenes were from SL3, Orbit 14, 5 August 1973, Tape 933847. Following these results, preparations were begun to process the multispectral scanner data using the program CANAL (canonical analysis) developed by the Office for Remote Sensing of Earth Resources (ORSER).[1] This procedure, with the subsequent analysis followed by a transformation of the data using the canonical transformation matrix generated by CANAL, was intended to accomplish the following:

1. improve the classification;
2. reduce the bulk of the data; and
3. increase our knowledge of the importance of the information contained in each band for classification of different targets.

PROCEDURE

The first step in the process was to obtain specific statistical information for each of the 22 signatures. This necessitated selection of permanent training areas. The training areas for each signature were defined by producing classification maps using each of the 22 original signatures as a category and then selecting the areas in which only that signature was mapped. This was accomplished for all but one signature (#15-urban), which occurred in blocks too small to be realistically defined as training areas.

The remaining permanent training areas defined as above for the 21 signatures were then processed by the program STATS. The following information was obtained for each of the signatures:

1. New signatures - These were obtained from the statistics of the permanent training areas. As was expected, these varied very little from the original signatures used in the classification maps from which the training areas were selected.

---

[1] See ORSER-SSEL Technical Reports 10-75 or 11-75 for descriptions of ORSER programs.

2. Histograms for each band - These were to check for a normal distribution of each signature and to insure that the training areas were correctly defined.

3. Variance-covariance matrix of mean result of each channel - This is an essential input for CANAL.

4. Number of observations in each training area - gain essential for CANAL.

Once the signatures, number of observations, and variance-covariance matrix for each of the signatures were completed they were used as input to the CANAL program.

Canonical, or multiple discriminant, analysis generates a linear transformation matrix C*. The transformation is of the form

$$Y_{ij} = C^*X_{ij}$$

where, in this case, X_{ij} is a sixteen channel raw Skylab element and Y_{ij} is a three axis transformed element. The object is to maximize the separability of the categories. This is done by maximizing the variance among categories subject to the constraint that the axes be orthogonal.

In addition, the transformed elements must have unit variance within categories. This may or may not reduce the variance within the category relative to the untransformed data. For Skylab and most other remotely sensed data, the transformation will usually reduce within-category variance. To the extent that it occurs, this reduction of within-category variance, coupled with the theoretically predicted maximum in among-category variance, constitutes a partial classification of the data.

RESULTS

The original transformation matrix, C, had sixteen axes. However, the first three axes contained 98.83% of the variance contained within the transformation. The values for axis one, two, and three were 83.61%, 14.49%, and 0.72%, respectively. Therefore, only the first three axes were used in transformation C*. Used in this way, the canonical transformation is a form of feature selection, since it means that data having sixteen spectral values for each pixel is transformed to data having only three values for each pixel. The result was an 81.25% reduction in data bulk. It is expected that using transformed data for classification will result in significant reductions in computer cost.

Table 1 lists the axis-to-channel correlation for the first three axes from the CANAL run. As can be seen, the highest correlations for axis 1 occur for channels 8, 9, 10, 19, 20, 12, and 17. Axis 2 (orthogonal to axis 1) has the highest axis-to-channel correlation for channels 1, 2, 3, 4, 5, 6, 12, 13, and 14. Finally, axis 3 (orthogonal to both axis 1 and axis 2) is highly correlated with channel 21 only. Although this axis contained only 0.72% of the total variance, it was decided to include it in the transformation simply because of its high correlation to the single thermal band.

Table 1: Axis-to-Channel Correlations

Orig. Ch. Number	Axes 1	2	3	Band Range, μm
1	0.07	0.48*	0.07	0.52- 0.56
2	0.07	0.48*	0.03	0.52- 0.56
3	0.07	0.46*	0.05	0.56- 0.61
4	0.07	0.46*	0.03	0.56- 0.61
5	0.06	0.43*	0.00	0.62- 0.67
6	0.06	0.44*	-0.00	0.62- 0.67
8	0.42*	0.26	0.05	0.68- 0.76
9	0.67*	-0.32	0.17	0.78- 0.88
10	0.61*	-0.29	0.17	0.78- 0.88
19	0.62*	-0.21	0.04	0.98- 1.03
20	0.75*	-0.18	-0.02	1.09- 1.19
12	0.49*	0.62*	-0.29	1.55- 1.75
13	0.20	0.60*	0.06	2.10- 2.35
14	0.20	0.60*	0.00	2.10- 2.35
21	0.01	0.35	0.85*	10.20-12.50
17	0.60*	-0.03	-0.13	12.00-13.00

*Strongest correlations of axis to channel.

Table 2: The Transformation Matrix, C*

Orig. Ch. Number	Axes 1	2	3
1	0.00131	0.01983	0.01179
2	-0.00189	0.03632	-0.01476
3	0.00292	0.03485	-0.02563
4	-0.01018	-0.00775	0.02719
5	-0.00248	0.02372	-0.02036
6	-0.00186	0.02269	-0.00213
8	0.02901	0.01816	0.01743
9	0.03093	-0.03286	0.02819
10	0.04762	-0.05659	0.03600
19	0.04222	-0.01758	0.03424
20	0.05387	-0.00378	-0.02608
12	0.05819	0.09958	-0.10537
13	0.02010	0.04062	0.01979
14	0.00182	0.03217	0.02168
21	0.00205	0.03708	0.15025
17	0.02898	0.00589	-0.02533

Table 2 lists the transformation matrix C* used in the transformation formula to transform the sixteen channels into three axes.

Having developed the transformation matrix and analyzed the resulting correlations, work to evaluate the classification, now possible using the transformed data, has begun. Hopefully, by knowing which channels are highly correlated to each axis and by observing the ability of the individual axis to map certain targets, we should get a better understanding of the importance of each individual band in mapping various target types.

ORSER-SSEL Technical Report 18-75
The Pennsylvania State University
August 1975